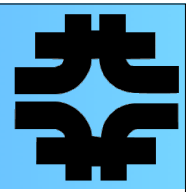


Computing Division CMS Meeting

Introduction and Overview — LATBauerdick
Facilities Projects and Operations — Jon Bakken
CMS Projects in CD I — Ian Fisk
CMS Projects in CD II — Greg Graham

Goals and Format of the Meeting

Information exchange -- give an overview of all work related to CMS

- in the division and beyond, as relevant to CD

Oversight and aligning activities to harmonize CD program for CMS

- all computing division activities related to CMS

Format of the meeting:

- series of relatively short talks covering CMS-related projects
- after initial meeting specific focus areas for each meeting
- contributions from all departments welcome — please propose
- avoid too much overlap with GDM meeting (unless deemed useful)
- lots of time for discussions — ~1h of prepared talks
- follow-up in new weekly department heads meeting or other venues

CMS Milestones

2004

- Mar/Apr. — DC04 to study T0 Reconstruction, Data Distribution, Real time analysis 25% of startup scale
- April — US-LCG Interoperability for job-submission (late)
- May — requirements and work plan for US Grid
- May/Jul — Data available and useable by PRS groups
- June — Data Streaming Project Demonstration (delayed to start in July)
- July — US work plan for distributed analysis environment
- Sep — PRS analysis feed-backs
- Sep — Draft CMS Computing Model in CHEP papers
- Nov — CMS/ARDA prototypes
- Nov — Data Analysis between Grids
- Dec — Computing TDR in initial draft form.

2005

- July — CMS Computing TDR released
- Post July?... — DC05 , 50% of startup scale.
- Dec — Physics TDR

2006

- DC06 — CMS final readiness tests, LCG TDR
- Fall. Computing Systems in place for LHC startup Continuous testing and preparations for data

CMS after DC04

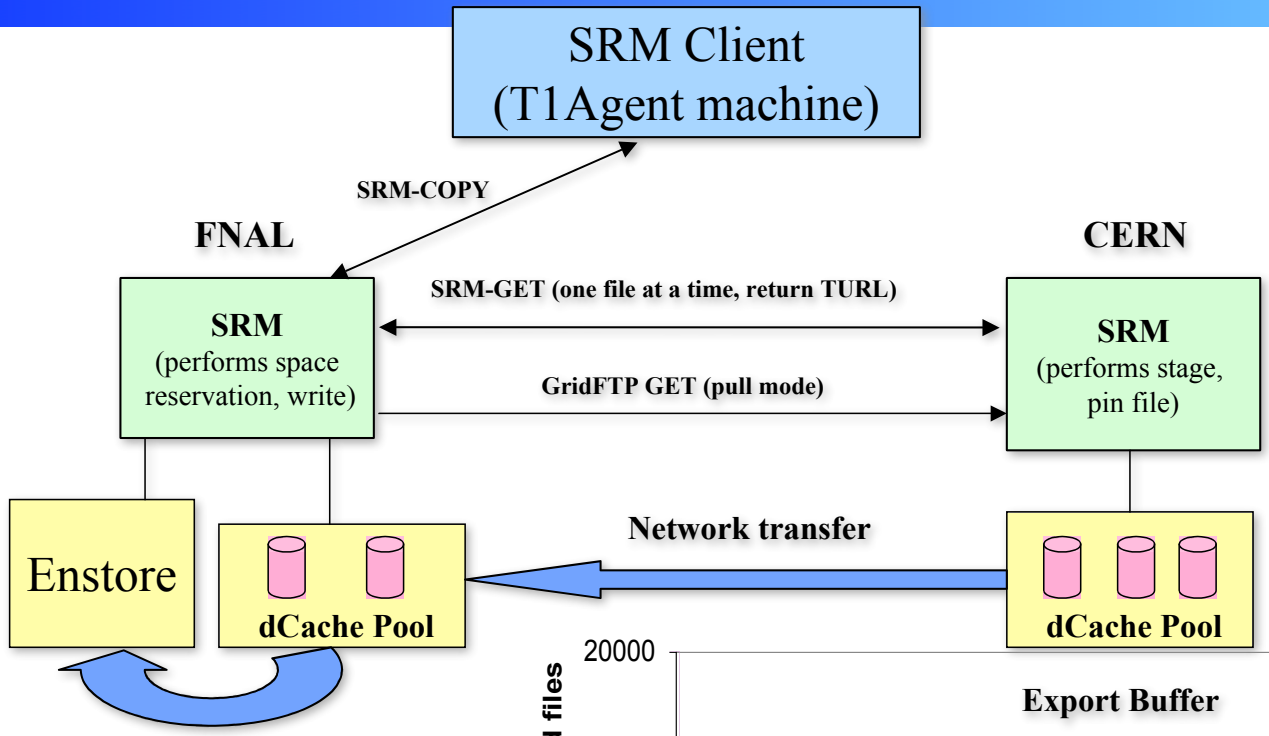
CMS Major Data Challenge DC04 in March/April 2004

- Many aims of DC04 achieved, but even more open questions
- focus on reconstruction program and DST at CERN, data transfers and handling, analysis at Tier-1/Tier-2 centers
  - reach a sustained 25Hz reconstruction rate in the Tier-0 farm (25% of the target conditions for LHC startup
  - register data and metadata to a catalogue
  - transfer the reconstructed data to all Tier-1 centers
  - analyze the reconstructed data at the Tier-1's as they arrive
  - publicize to the community the data produced at Tier-1's
  - monitor and archive of performance criteria of the ensemble of activities for debugging and post-mortem analysis

Pre-challenge production of 2003/04 is continuing, and being sustained

- 70M Monte Carlo events (30M with Geant-4)
- US CMS contribution of Geant4 production entirely done on Grid3

Fermilab SRM for DC04 File Transfers



Total data transferred to FNAL

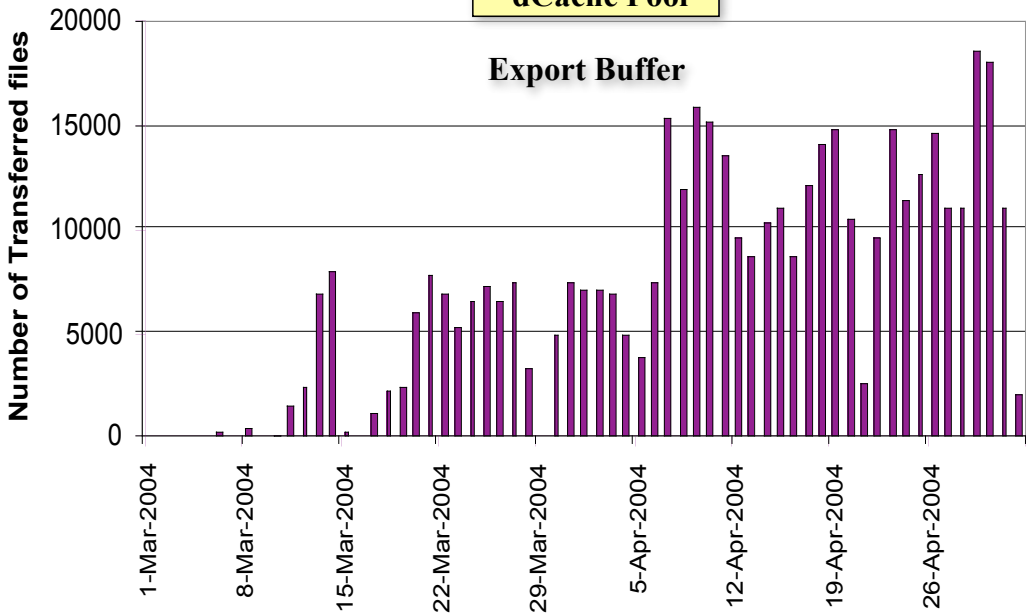
- 5.2TB (5293GB)

Total number of files transferred:

- 440K

Average filesize very small:

- median: 581.6KB



After DC04: Data Handling Top Priority

Expect significant shift of expectation on S&C for CMS physics preparations

- development project gets strong "(US) CMS offline group" component
- provide valid and useable software environment and operations

Focussed CMS-Prototype for "Distributed Analysis" end-to-end system

- make data available to CMS physicists for "offline analysis"
- address data handling, data management, data access issues

application level services: handling / management of CMS data sets and jobs

- data set services
- catalogs: file metadata, replica, event and data set meta data
- distributed file delivery and caching services
- storage services: permanent and transient, managed and reliable
- management and tracking of processing/analysis jobs

provide and further develop the required fabric resources

- farms and compute services, including "opportunistic" resources
- disks/tapes and storage services, data movers and data access
- VO management and AAA services, security and policy infrastructure

Current Projects: CMS Software

Many things going on with Fermilab involvement in and outside of CD-CMS

- Framework developments and POOL
- CMS Event Data Model: DST workshop and LPC review group
- Investigations on using ROOT on CMS data
- Mathlib project with LCG
- Software Development Environment: LPC review group
- HCAL Database -- Briefing and Lee's talk at CERN, FroNTier
- Metadata and Data Management: SAMgrid, rcp
- LPC Software Support -- trying to understand and define scope

not covered today!

Current Projects: US User Facilities

See talks by Jon, Ian, Greg

- Preparation of CMS datasets at Fermi and data set service
- CMS MC production at 10M events/month
- Interoperability with LCG
- VOMRS and populating our VOMS with LCG,
- Privilege/authorization project
- Storage Services, SRM/dCache
- Robust Data Transfers
- Runjob for CMS production and analysis
- MC generation service for individuals
- Test of file catalogs and data management components
  SAM, Globus RLS, CMS/ARDA file catalog

Tier-2 Meeting at Caltech

"Advisory S&C Board Retreat" on US CMS Tier-2 Centers

- The goal of the meeting is to help define the scope and goals for the US CMS Tier-2 work plan, and to set the expectations and requirements for CMS Tier-2 centers in the US
- The result of the meeting will be a document advising the US CMS Software and Computing project manager on starting the Tier-2 program for US CMS, and will be input in the bidding procedure for selecting the US CMS Tier-2 sites.
- Discuss roles and responsibilities of US Tier-2 centers for providing part of the computing fabric, production services, analysis services
- Outcome relevant to Fermilab regarding role of the Tier-1 to support Tier-2 centers and as a "portal" to LCG

Mini-Review in July

Review meeting July 8 at the DOE

received set of questions ~ as expected

- on Management, Software, Facilities&Grids
- new focus on "Physics Analysis Model"
  - What are your plans for a physics analysis center?
  - Do your plans for a physics analysis center affect the role of (any) Tier-2 centers? If so, in what ways?
  - What is the status of your physics analysis effort and plans for the future?
    - How does US-(ATLAS,CMS) compare to the rest of the collaboration in terms of physics preparedness?
    - How do you plan on tracking progress in this area? When/how do you validate the model?

New Projects: Data Management

Requirements and Technical Assessment Group in CMS

- strong Fermilab particip.: Rob H. ed. , Avi Y. on PRS side, LATB chair

Draft charge and mandate for the CPT Requirements and Technical Assessment Group (RTAG) on a CMS Data Management System (DMS)

After the CMS data challenge DC04, CPT needs to define the required functionalities and do a technical assessment of the CMS Data Management System (DMS), in order to define a sub-task in CCS to engineer and implement such a system. The CPT Joint Technical Board thus is starting an RTAG to address these issues.

The requirements for the CMS DMS are defined by the needs of the PRS groups and the use cases for performing data analysis, event simulation and data handling by physicists, the CMS offline group and the regional center staff in the LHC distributed computing environment. The DMS should be build on top of the emerging grid tools and grid services and interface to the CMS COBRA framework and the (yet to be defined) environment for Distributed Analysis (DA). A blueprint for the CMS DMS needs to be developed in time for the CCS Computing TDR.

Motivation for the RTAG:
-- to agree on a set of use cases to be addressed by the CMS data management system -- to agree on requirements which will allow CCS to provide a focus of effort
-- to provide guidance to CCS on development directions and interfacing work to match the requirements, defining the scope of the CMS data management task
-- to clearly identify the roles and responsibilities of the components/layers/services in CMS data management, distributed analysis and LHC computing grid services
-- To give guidance to the community on the expected division of work between the experiments, the LCG and the external projects.

Mandate:
-- review, reconcile and define the use cases and CMS requirements for Data Management and capture them in a consistent way, taking into account the agreed HEPCAL use cases and the existing and expected sets of middlewares, experience from other experiemnts, the CMS application environment and application services like COBRA, POOL, the CMS production environment and the CMS user's potential work environments etc.
-- consider the interfaces of a CMS DMS to Cobra/POOL, the distributed analysis environment, grid middleware and LHC computing grid services and experiment-specific services
-- consider the functionality of existing CMS packages, state of advancement and role in the experiment and identify functionalities and components that could be integrated in the CMS DMS
-- develop a roadmap specifying wherever possible the architecture, the components and potential sources of deliverables to guide the medium term work of CCS and the DMS and DA planning in the experiment

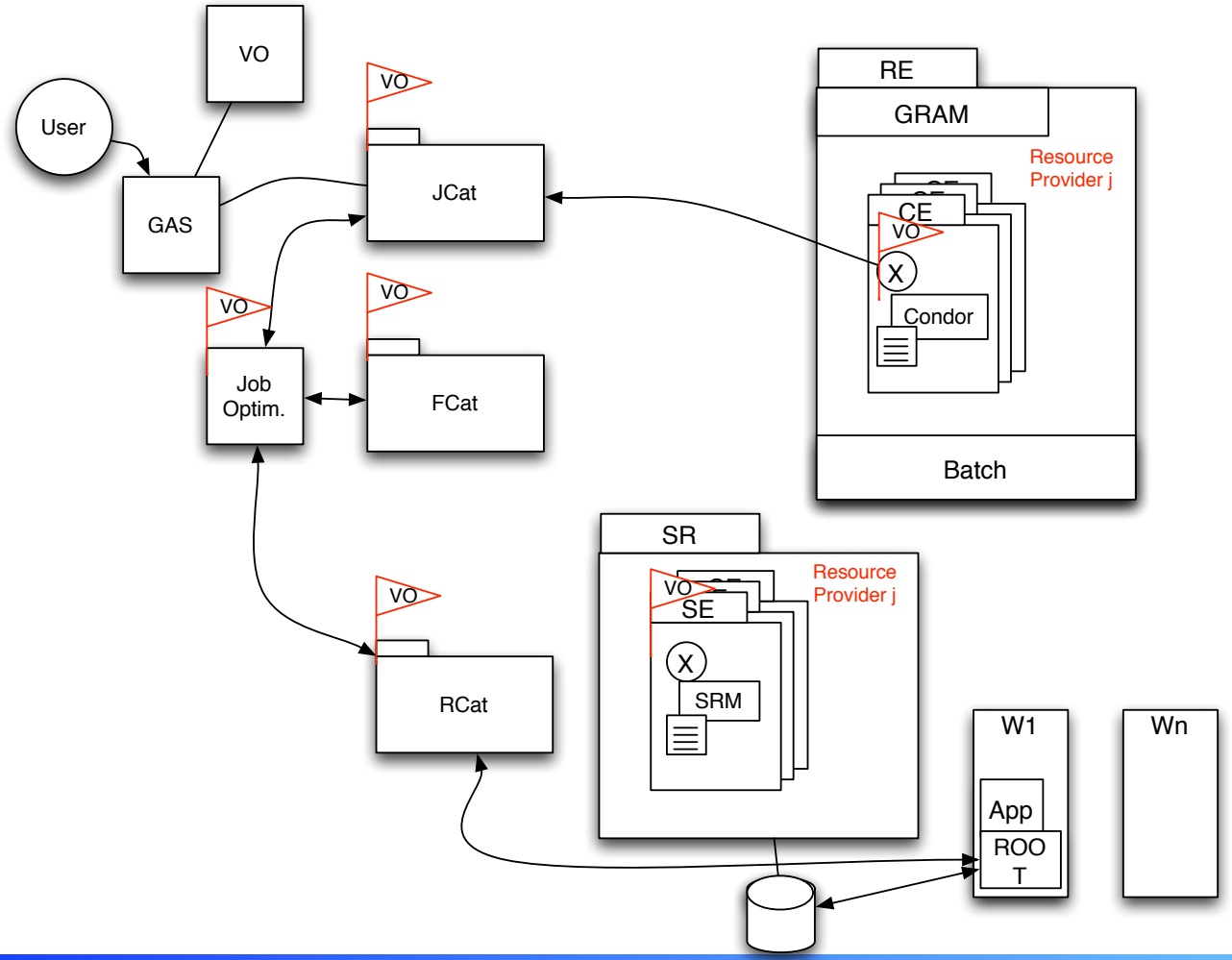
Schedule:
The CMS DMS RTAG shall provide a draft report to the CPT JTB by August 2004. It should contain initial guidance to inform the CPT Joint Technical Board about a program of work and sub-task to implement the CMS DMS.

The final RTAG report is expected for September 2004.

More New Projects: CMS/ARDA

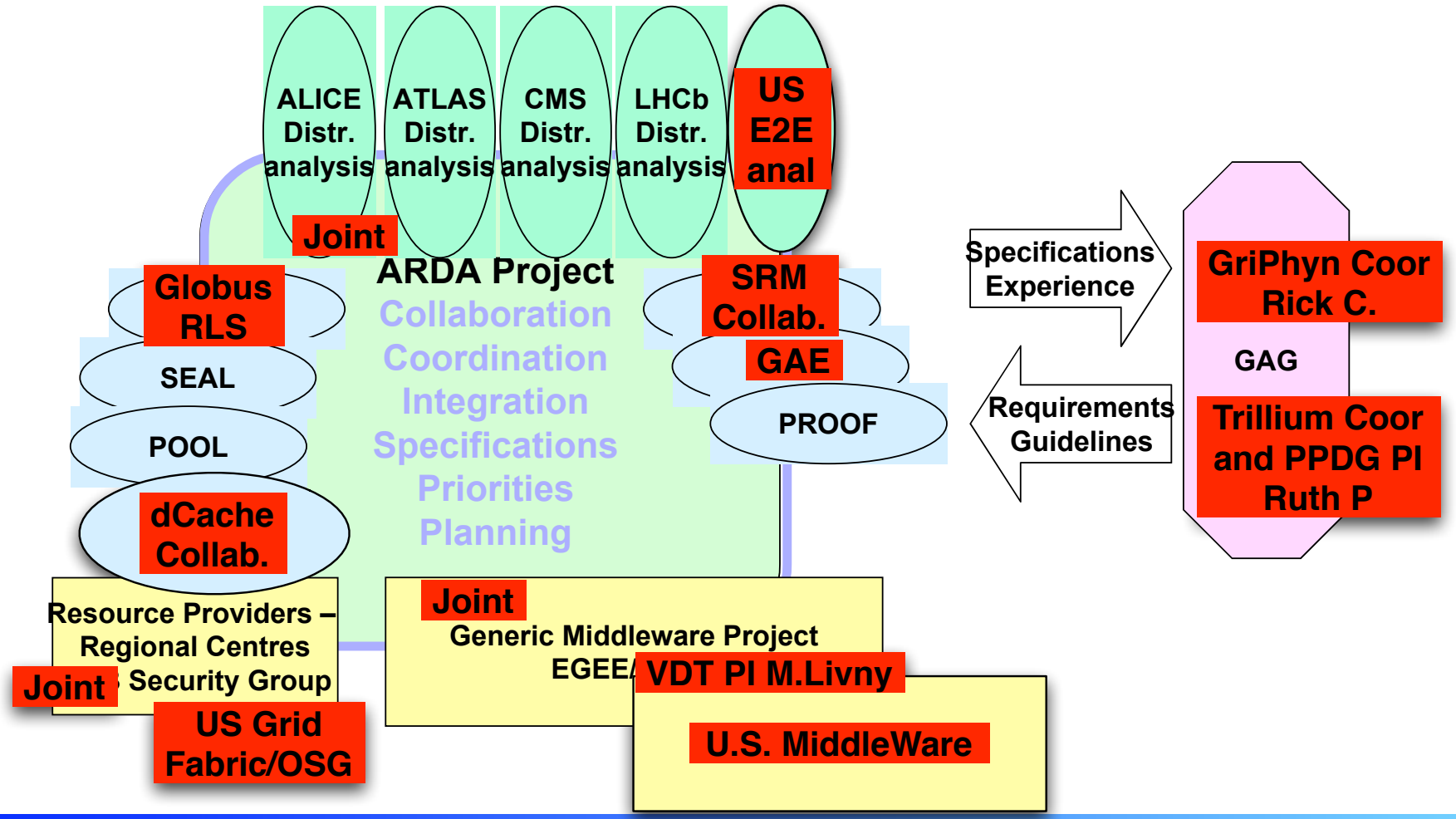
new CMS project for distributed analysis and the new round of middleware

"gLite" prototype released, plans to look at it and get some experience



ARDA and US/Fermilab?

Can and should the US and Fermilab get involved in the ARDA project?



Working with others

Joint Steering Group and joint USCMS-OSG projects and milestones

Model of working together:

- agreed milestones and programs of work
- well-defined projects within this larger context

Initial set of Joint Projects being defined

- sustaining the infrastructure: Grid3 --> Open Science Grid
- Open Science Grid storage services working group
- Joint Security working group: LCG and US security model and work
- Open Science Grid Blueprint for Application Grid Services
- LCG Interoperability and LCG Service Challenges
  - initial project: sustained Data Transfers b/w CERN and Tier-1 centers
  - Fermilab-CERN part done as a CD-IT project in the context of the OSG "Data Management and Replication Challenge"

Start of the Monthly CMS Meeting

Overview on current program of work on US User Facilities at Fermilab

- Facilities Projects and Operations — Jon Bakken
- CMS Projects in CD I — Ian Fisk
- CMS Projects in CD II — Greg Graham

next week: work with CCF on "harmonizing" CCF and CMS WBS